ガウス過程モデルと Random Fourier Features

石川 徹也　<tiskw111@gmail.com>

2021 年 10 月 13 日

はじめに

本文書では，ガウス過程モデル [1] に random Fourier features [2] を適用する手順を解説します．これにより，ガウス過程モデルの学習や推論を高速化させることができ，より大規模なデータにガウス過程モデルを適用することができます．

ガウス過程モデル [1] は確率的教師あり機械学習フレームワークのひとつであり，サポートベクトルマシンやランダムフォレストなどと並んで，回帰や分類タスクに広く使用されています．ガウス過程モデルがサポートベクトルマシンやランダムフォレストと大きく違う点は「確率的なモデルである」ことです．すなわち，ガウス過程モデルは確率的なモデルとして定式化されているため，予測値だけでなく，その予測に対する不確実性の尺度をも提供することができます．これは機械学習の説明性を上げることのできる非常に有益な性質です．

その一方で，ガウス過程モデルは学習や推論の計算コストが高いことでも知られています．学習データの総数を $N \in \mathbb{Z}^+$ としたとき，ガウス過程モデルの学習に要する計算量は $O(N^3)$，推論に要する計算量は $O(N^2)$ です．問題は計算量が学習データの総数 N のべき乗になってしまっていることで，これは大規模データへの適用に際して障害になり得ます．これはガウス過程モデルがカーネル法と同等の数学的構造を有していることに起因しており，言い変えれば，カーネルサポートベクトルマシンも同じ悩みを有しています．

カーネル法を高速化する手法のひとつに random Fourier features [2] があります（以下 RFF と略します）．これはカーネル関数を有限次元ベクトルの内積として近似することで，カーネル法の柔軟性を維持しつつ計算量を大幅に削減する手法です．具体的には学習に要する計算量を $O(ND^2)$，推論に要する計算量を $O(D^2)$ にまで削減することができます．ただし $D \in \mathbb{Z}^+$ は RFF のハイパーパラメータであり，学習データの総数 N とは独立に指定することができます．

ガウス過程モデルはカーネル法と同等の数学的構造を有しているため，ガウス過程モデルにも RFF を適用することができます．これにより，ガウス過程モデルはより強力かつお手軽に使える，非常に頼もしいツールへと進化します．

しかしながら，ガウス過程モデルへの RFF の適用にあたっては，実は一筋縄ではいかないところがあります．ただ RFF を適用するだけでは高速化につながらず，一工夫する必要があるのです．ですが，そのあたりの困難さや解決方法に言及している文献が，残念ながら世の中には存在しないようでしたので，本文書にてその手順を解説しようと考えた次第です．

ちなみに，本文書は拙作ライブラリ `rfflearn` [4] に同梱されているドキュメントの日本語版です．おそらく `rfflearn` に同梱されている文書の方が頻繁にメンテナンスされると思いますので，英語でも差し支えない方はそちらをご参照下さい．

NOTE: 上述のライブラリ `rfflearn` は以下で公開されています．

`https://github.com/tiskw/random-fourier-features`

1　ガウス過程モデル再訪

本節ではガウス過程モデルの概要について述べます．残念ながら本文書ではガウス過程モデルの定式化や導出などの詳細は扱いませんので，詳細にご興味のある読者は Rasmussen [2] あるいは赤穂 [3] をご参照下さい．

学習データを $\mathcal{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^N$，ラベルの観測誤差の標準偏差を $\sigma \in \mathbb{R}^+$ とします．ただし $\boldsymbol{x}_n \in \mathbb{R}^M$, $y_n \in \mathbb{R}$ とします．このときガウス過程モデルは，テストデータ $\boldsymbol{\xi} \in \mathbb{R}^M$ の予測値の期待値を

$$m(\boldsymbol{\xi}) = \widehat{m}(\boldsymbol{\xi}) + (\boldsymbol{y} - \widehat{\boldsymbol{m}})^\mathsf{T} \left(\boldsymbol{K} + \sigma^2 \boldsymbol{I}\right)^{-1} \boldsymbol{k}(\boldsymbol{\xi}), \qquad (1)$$

で与え，さらにテストデータ $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ の予測値の共分散を

$$v(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = k(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) - \boldsymbol{k}(\boldsymbol{\xi}_1)^\mathsf{T} \left(\boldsymbol{K} - \sigma^2 \boldsymbol{I}\right)^{-1} \boldsymbol{k}(\boldsymbol{\xi}_2), \qquad (2)$$

で与えます．ただし関数 $k : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ はカーネル関数，行列 $\boldsymbol{K} \in \mathbb{R}^{N \times N}$ は

$$\boldsymbol{K} = \begin{pmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_N, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{pmatrix}, \qquad (3)$$

で与えられる行列，ベクトル $\boldsymbol{k}(\boldsymbol{\xi}) \in \mathbb{R}^N$ は

$$\boldsymbol{k}(\boldsymbol{\xi}) = \begin{pmatrix} k(\boldsymbol{\xi}, \boldsymbol{x}_1) \\ \vdots \\ k(\boldsymbol{\xi}, \boldsymbol{x}_N) \end{pmatrix}, \qquad (4)$$

です．ベクトル $\boldsymbol{y} \in \mathbb{R}^N$ は学習データのラベルをベクトル状に並べたものであり，$\boldsymbol{y} = (y_1, y_2, \ldots, y_N)^\mathsf{T}$ です．また $\widehat{m}(\boldsymbol{\xi})$ は予測値の事前分布であり，$\widehat{\boldsymbol{m}}$ は学習データの予測値の事前分布をベクトル状に並べたものです．特に事前分布を設定する必要がない場合は $\widehat{m}(\cdot) = 0$, $\widehat{\boldsymbol{m}} = \boldsymbol{0}$ とするのが一般的です．

テストデータ $\boldsymbol{\xi}$ の予測値の分散を求めるためには，共分散を求める式 (2) において $\boldsymbol{\xi}_1 = \boldsymbol{\xi}_2 = \boldsymbol{\xi}$ とすれば良く，

$$v(\boldsymbol{\xi},\boldsymbol{\xi}) = k(\boldsymbol{\xi},\boldsymbol{\xi}) - \boldsymbol{k}(\boldsymbol{\xi})^\mathsf{T}\left(\boldsymbol{K} - \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{k}(\boldsymbol{\xi}), \tag{5}$$

となります．

2 RFF 再訪

本節では RFF の概要について述べます．こちらも残念ながら詳細について述べる紙面的余裕がありませんので，詳細をお知りになりたい場合は原論文 [1] をご参照下さい．

関数 $k : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$ をカーネル関数とします．カーネル関数 k を Fourier 変換することで

$$k(\boldsymbol{x}_1,\boldsymbol{x}_2) \simeq \boldsymbol{\phi}(\boldsymbol{x}_1)^\mathsf{T}\boldsymbol{\phi}(\boldsymbol{x}_2), \tag{6}$$

という近似式を求めるのが RFF です．このとき右辺のベクトル $\boldsymbol{\phi}(\boldsymbol{x}_1)$ の次元 D は任意に設定することができ，次元 D が大きいほど式 (6) の近似精度が高い一方で，次元 D を大きくすると計算量が多くなります．

具体例をひとつ挙げておきましょう．カーネル関数のひとつとして有名な RBF カーネルは

$$k(\boldsymbol{x}_1,\boldsymbol{x}_2) = \exp\left(-\gamma\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2\right), \tag{7}$$

と与えられます．この RBF カーネルに対して RFF を適用すると

$$\boldsymbol{\phi}(\boldsymbol{x}) = \begin{pmatrix}\cos \boldsymbol{W}\boldsymbol{x} \\ \sin \boldsymbol{W}\boldsymbol{x}\end{pmatrix}, \tag{8}$$

となります．ただし行列 $\boldsymbol{W} \in \mathbb{R}^{D/2\times M}$ は，各要素を正規分布 $\mathcal{N}(0, \frac{1}{4\gamma})$ にしたがってサンプリングしたランダム行列です．

3 ガウス過程モデルと RFF

本節ではガウス過程モデルに RFF を適用し，高速化の効果を理論的に確認します．

3.1 RFF 適用前のガウス過程モデルの計算量

まずは通常のガウス過程モデルの学習および推論に要する計算量を確認しておきましょう．前提として，入力ベクトルの次元 M よりも学習データ数 N の方が十分に大きいと仮定します．このとき式 (1) および (2) のうち, 学習時間のボトルネックは明らかに逆行列 $\left(\boldsymbol{K} + \sigma^2\boldsymbol{I}\right)^{-1}$ の計算にあります．この行列の大きさは $N \times N$ ですので，学習に要する計算量は $O(N^3)$ となります．次に推論ですが，テスト時のボトルネックは行列積 $(\boldsymbol{y} - \widehat{\boldsymbol{m}})^\mathsf{T}\left(\boldsymbol{K} + \sigma^2\boldsymbol{I}\right)^{-1}$ あるいは $\boldsymbol{k}(\boldsymbol{\xi}_1)^\mathsf{T}\left(\boldsymbol{K} - \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{k}(\boldsymbol{\xi}_2)$ であり，これらの行列積に要する計算量はいずれも $O(N)$ となります．

3.2 予測値の期待値への RFF の適用

さて，ではいよいよガウス過程モデルに RFF を適用します．まずは予測値の期待値ですが，式 (1) に RFF の近似式 (2) を代入すると

$$m(\boldsymbol{\xi}) = \widehat{m}(\boldsymbol{\xi}) + (\boldsymbol{y} - \widehat{\boldsymbol{m}})^\mathsf{T}\left(\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi} + \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\phi}(\boldsymbol{\xi}), \tag{9}$$

となります．ただし行列 $\boldsymbol{\Phi}$ は RFF によって得られるベクトル $\boldsymbol{\phi}$ を学習データ全てに対して並べた $D \times N$ 行列 $\boldsymbol{\Phi} = (\boldsymbol{\phi}(\boldsymbol{x}_1), \ldots, \boldsymbol{\phi}(\boldsymbol{x}_N))$ です．しかし，まだこれでは高速化は図れていません．式 (9) の計算量のボトルネックは依然として $N \times N$ 行列の逆行列のままです．

ここで一工夫します．式 (9) に対して逆行列の反転補題 (*binominal inverse lemma*) を適用することを考えます．逆行列の反転補題とは以下の定理です．

定理 3.1 (逆行列の反転補題)

行列 $\boldsymbol{A} \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$, $\boldsymbol{B} \in \mathbb{R}^{\mathbb{N}\times\mathbb{M}}$, $\boldsymbol{C} \in \mathbb{R}^{\mathbb{M}\times\mathbb{N}}$, $\boldsymbol{D} \in \mathbb{R}^{\mathbb{M}\times\mathbb{M}}$ に対して以下が成り立つ．

$$\begin{aligned}&(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{D}\boldsymbol{C})^{-1} \\ &\quad = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D}^{-1} + \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}\end{aligned} \tag{10}$$

ただし行列 $\boldsymbol{A}$, $\boldsymbol{D}$ は正則行列とする．

証明は本文書の末尾で行うものとし，ここではガウス過程モデルへの RFF の適用の話を先に進めさせて下さい．上記の補題に対して $\boldsymbol{A} = \sigma^2\boldsymbol{I}$, $\boldsymbol{B} = \boldsymbol{\Phi}^\mathsf{T}$, $\boldsymbol{C} = \boldsymbol{\Phi}$, $\boldsymbol{D} = \boldsymbol{I}$ とおけば,

$$\left(\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi} + \sigma^2\boldsymbol{I}\right)^{-1} = \frac{1}{\sigma^2}\left(\boldsymbol{I} - \boldsymbol{\Phi}^\mathsf{T}\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^\mathsf{T} + \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{\Phi}\right),$$

を得ます．ここでさらに $\boldsymbol{P} = \boldsymbol{\Phi}\boldsymbol{\Phi}^\mathsf{T} \in \mathbb{R}^{D\times D}$ とおき，上式の両辺に右から $\boldsymbol{\Phi}$ をかけると

$$\left(\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi} + \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{\Phi} = \frac{1}{\sigma^2}\boldsymbol{\Phi}^\mathsf{T}\left(\boldsymbol{I} - \left(\boldsymbol{P} + \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{P}\right), \tag{11}$$

を得ます．したがって式 (9) は

$$m(\boldsymbol{\xi}) = \widehat{m}(\boldsymbol{\xi}) + \frac{1}{\sigma^2}(\boldsymbol{y} - \widehat{\boldsymbol{m}})^\mathsf{T}\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{S}, \tag{12}$$

と書き改めることができます．ただし

$$\boldsymbol{S} = \boldsymbol{I} - \left(\boldsymbol{P} + \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{P}, \tag{13}$$

です．

聡明なる読者は，すでにボトルネックが解消されていることにお気付きのことと思います．式 (9) のボトルネックであった逆行列 $(\boldsymbol{K} + \sigma^2\boldsymbol{I})^{-1}$ は，式 (12), (13) では $(\boldsymbol{P} + \sigma^2\boldsymbol{I})^{-1}$ となり，行列のサイズは $D \times D$ になりました．通常，RFF の次元 D は学習データの総数 N よりも十分小さく設定しますので，もはやこの逆行列の計算はボトルネックではなくなりました．式 (12), (13) のボトルネックは行列積 $\boldsymbol{P} = \boldsymbol{\Phi}\boldsymbol{\Phi}^\mathsf{T}$ であり，この計算量は $O(ND^2)$ です．元のガウス過程モデルの学習に要する計算量が $O(N^3)$ であったことを振り返れば，RFF によってかなりの高速化が達成できたことになります．

3.3 予測値の共分散への RFF の適用

次に予測値の共分散 (2) に RFF を適用していきましょう．式 (2) に対して RFF の近似式 (2) を代入し，さらに式 (11) を適用すると

$$v(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) = \boldsymbol{\phi}(\boldsymbol{\xi}_1)^\mathsf{T}\boldsymbol{\phi}(\boldsymbol{\xi}_2) - \frac{1}{\sigma^2}\boldsymbol{\phi}(\boldsymbol{\xi}_1)^\mathsf{T}\boldsymbol{P}\boldsymbol{S}\boldsymbol{\phi}(\boldsymbol{\xi}_2)$$

Algorithm 1: RFF 適用後のガウス過程モデルの学習

Data: $\mathcal{D} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^N$, $\sigma \in \mathbb{R}^+$
Result: $\boldsymbol{c}_\mathrm{m} \in \mathbb{R}^D$, $\boldsymbol{C}_\mathrm{v} \in \mathbb{R}^{D\times D}$
$\boldsymbol{y} \leftarrow (y_1, \ldots, y_N)^\mathsf{T}$
$\boldsymbol{\Phi} \leftarrow (\boldsymbol{\phi}(\boldsymbol{x}_1), \ldots, \boldsymbol{\phi}(\boldsymbol{x}_N))$
$\boldsymbol{P} \leftarrow \boldsymbol{\Phi}\boldsymbol{\Phi}^\mathsf{T}$
$\boldsymbol{S} \leftarrow \boldsymbol{I} - \left(\boldsymbol{P} + \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{P}$
$\boldsymbol{c}_\mathrm{m} \leftarrow \frac{1}{\sigma^2}\boldsymbol{y}^\mathsf{T}\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{S}$ /* 予測値の期待値の算出に使用 */
$\boldsymbol{C}_\mathrm{v} \leftarrow \boldsymbol{I} - \frac{1}{\sigma^2}\boldsymbol{P}\boldsymbol{S}$ /* 予測値の共分散の算出に使用 */

Algorithm 2: RFF 適用後のガウス過程モデルの推論

Data: $\boldsymbol{\xi} \in \mathbb{R}^M$, $\boldsymbol{c}_\mathrm{m} \in \mathbb{R}^D$, $\boldsymbol{C}_\mathrm{v} \in \mathbb{R}^{D\times D}$
Result: $\mu \in \mathbb{R}$, $\eta \in \mathbb{R}$
$\boldsymbol{z} \leftarrow \boldsymbol{\phi}(\boldsymbol{\xi})$
$\mu \leftarrow \boldsymbol{c}_\mathrm{m}\boldsymbol{z}$ /* 予測値の期待値の算出 */
$\eta \leftarrow \boldsymbol{z}^\mathsf{T}\boldsymbol{C}_\mathrm{v}\boldsymbol{z}$ /* 予測値の共分散の算出 */

表 1　RFF 適用前後でのガウス過程モデルの計算量

	学習	推論
RFF 適用前	$O(N^3)$	$O(N)$
RFF 適用後	$O(ND^2)$	$O(D^2)$

$$= \boldsymbol{\phi}(\boldsymbol{\xi}_1)^\mathsf{T}\left(\boldsymbol{I} - \frac{1}{\sigma^2}\boldsymbol{P}\boldsymbol{S}\right)\boldsymbol{\phi}(\boldsymbol{\xi}_2), \tag{14}$$

となります．式 (14) のボトルネックは，予測値の期待値と同様に行列積 $\boldsymbol{P} = \boldsymbol{\Phi}\boldsymbol{\Phi}^\mathsf{T}$ であり，この計算量は $O(ND^2)$ です．

ここで，RFF を適用した後のガウス過程モデルの学習および推論の手順を疑似コードとして Algorithm 1, 2 にまとめておきます．ただし Algorithm 1, 2 では簡単のために事前分布を 0 としています．

最後に，RFF を適用した後の計算量を表 1 に整理しました．ただし $N \in \mathbb{Z}^+$ は学習データの総数，$D \in \mathbb{Z}^+$ は RFF の次元です．

付録 A　補足

A.1　逆行列の反転補題の証明

逆行列の反転補題を再掲し，証明します．

定理 付録 A.1 (逆行列の反転補題)

行列 $\boldsymbol{A} \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$, $\boldsymbol{B} \in \mathbb{R}^{\mathbb{N}\times\mathbb{M}}$, $\boldsymbol{C} \in \mathbb{R}^{\mathbb{M}\times\mathbb{N}}$, $\boldsymbol{D} \in \mathbb{R}^{\mathbb{M}\times\mathbb{M}}$ に対して以下が成り立つ．

$$(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{D}\boldsymbol{C})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D}^{-1} + \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}$$

ただし行列 $\boldsymbol{A}$, $\boldsymbol{D}$ は正則行列とする．

証明： 以下の等式が成立する．

$$\begin{aligned}\begin{pmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{pmatrix}^{-1} &= \begin{pmatrix}\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S}\\ -\boldsymbol{S}\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{S}\end{pmatrix}\\ &= \begin{pmatrix}\boldsymbol{T} & -\boldsymbol{T}\boldsymbol{B}\boldsymbol{D}^{-1}\\ -\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{T} & \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{T}\boldsymbol{B}\boldsymbol{D}^{-1}\end{pmatrix},\end{aligned}$$

ただし

$$\boldsymbol{T} = \left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}, \tag{15}$$
$$\boldsymbol{S} = \left(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1}, \tag{16}$$

とする．これは直接計算により明らか．さて，上記ブロック行列の対応する箇所を比較することで

$$\boldsymbol{T} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S}\boldsymbol{C}\boldsymbol{A}^{-1}, \tag{17}$$
$$\boldsymbol{S} = \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{T}\boldsymbol{B}\boldsymbol{D}^{-1}, \tag{18}$$
$$-\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S} = -\boldsymbol{T}\boldsymbol{B}\boldsymbol{D}^{-1}, \tag{19}$$
$$-\boldsymbol{S}\boldsymbol{C}\boldsymbol{A}^{-1} = -\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{T}, \tag{20}$$

を得る．式 (17) に対して

$$\boldsymbol{A} \to \boldsymbol{D}^{-1}, \quad \boldsymbol{B} \to -\boldsymbol{C}, \quad \boldsymbol{C} \to \boldsymbol{B}, \quad \boldsymbol{D} \to \boldsymbol{A},$$

と置き直せば，証明すべき式を得る． ■

おわりに

本文書では，ガウス過程モデルへ RFF を適用する具体的な手順をご紹介し，理論的な計算量を導出しました．具体的な実装にご興味のある方は，拙作のライブラリ `rfflearn` [4] をご参照下さい．

最後に，私の数学的活動は，2017 年に逝去された恩師，山下弘一郎先生や，大学および大学院で私の指導教官を担当して下さった早川朋久准教授をはじめ，数学で私と関わりを持ったすべての方々のおかげで成り立っています．そして，数学的活動の以前に，そもそも私の生は両親によって与えられ，妻によって支えられています．

参考文献

[1] A. Rahimi and B. Recht, “Random Features for Large-Scale Kernel Machines”, Neural Information Processing Systems, 2007.
[2] C. Rasmussen and C. Williams, “Gaussian Processes for Machine Learning”, MIT Press, 2006.
[3] 赤穂昭太郎, “ガウス過程回帰の基礎”, システム/制御/情報, vol. 62, no. 10, pp. 390-395, 2018. https://www.jstage.jst.go.jp/article/isciesci/62/10/62_390/_pdf
[4] https://github.com/tiskw/random-fourier-features